

Model Optimization and Tuning Phase

Date	14 June 2025
Team ID	SWTID1749627644
Project Title	Human Resource Management: Predicting Employee Promotions using Machine Learning
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Decision Tree Classifier	<pre>def decisionTree_tuned(x_train, y_train, x_test, y_test): print('Tuned Decision Tree Classifier') params = { 'max_depth': [5, 10, None], 'min_samples_split': [2, 5, 10], 'criterion': ['gini', 'entropy'] }</pre>	<pre>Tuned Decision Tree Classifier Best Parameters: {'criterion': 'entropy', 'max_depth': None, 'min_samples_split': 2}</pre>
Random Forest Classifier	<pre>def randomForest_tuned(x_train, y_train, x_test, y_test): print('Tuned Random Forest Classifier') params = { 'n_estimators': [100, 200], 'max_depth': [10, None], 'min_samples_split': [2, 5], 'criterion': ['gini', 'entropy'] }</pre>	<pre>Tuned Random Forest Classifier Best Parameters: {'criterion': 'gini', 'max_depth': None, 'min_samples_split': 2, 'n_estimators': 100}</pre>

KNN Classifier	<pre>def KNN_tuned(x_train, y_train, x_test, y_test): print('Tuned KNN Classifier') params = { 'n_neighbors': [3, 5, 7], 'weights': ['uniform', 'distance'], 'p': [1, 2] # 1 = Manhattan, 2 = Euclidean } # ... (rest of the code) ...</pre>	<pre>Tuned KNN Classifier Best Parameters: {'n_neighbors': 3, 'p': 1, 'weights': 'distance'} Confusion Matrix:</pre>
XGBoost Classifier	<pre>def xgboost_tuned(x_train, y_train, x_test, y_test): print('Tuned XGBoost Classifier (sklearn version)') params = { 'n_estimators': [100, 200], 'learning_rate': [0.05, 0.1], 'max_depth': [3, 5], 'subsample': [0.8, 1.0] } # ... (rest of the code) ...</pre>	<pre>Tuned XGBoost Classifier (sklearn version) Best Parameters: {'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8}</pre>

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Decision Tree Classifier	<pre>Confusion Matrix: [[13892 1173] [886 14133]] Classification Report: precision recall f1-score support 0 0.94 0.92 0.93 15065 1 0.92 0.94 0.93 15019 accuracy 0.93 0.93 0.93 30084 macro avg 0.93 0.93 0.93 30084 weighted avg 0.93 0.93 0.93 30084</pre>

Random Forest Classifier	<div>Confusion Matrix: [[14183 882] [769 14250]] Classification Report:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.95</td><td>0.94</td><td>0.94</td><td>15065</td></tr><tr><td>1</td><td>0.94</td><td>0.95</td><td>0.95</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.95</td><td>30084</td></tr><tr><td>macro avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.95	0.94	0.94	15065	1	0.94	0.95	0.95	15019	accuracy			0.95	30084	macro avg	0.95	0.95	0.95	30084	weighted avg	0.95	0.95	0.95	30084
	precision	recall	f1-score	support																											
0	0.95	0.94	0.94	15065																											
1	0.94	0.95	0.95	15019																											
accuracy			0.95	30084																											
macro avg	0.95	0.95	0.95	30084																											
weighted avg	0.95	0.95	0.95	30084																											
KNN Classifier	<div>Confusion Matrix: [[12965 2100] [731 14288]] Classification Report:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.95</td><td>0.86</td><td>0.90</td><td>15065</td></tr><tr><td>1</td><td>0.87</td><td>0.95</td><td>0.91</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.91</td><td>30084</td></tr><tr><td>macro avg</td><td>0.91</td><td>0.91</td><td>0.91</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.91</td><td>0.91</td><td>0.91</td><td>30084</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.95	0.86	0.90	15065	1	0.87	0.95	0.91	15019	accuracy			0.91	30084	macro avg	0.91	0.91	0.91	30084	weighted avg	0.91	0.91	0.91	30084
	precision	recall	f1-score	support																											
0	0.95	0.86	0.90	15065																											
1	0.87	0.95	0.91	15019																											
accuracy			0.91	30084																											
macro avg	0.91	0.91	0.91	30084																											
weighted avg	0.91	0.91	0.91	30084																											

XGBoost Classifier	<pre> Confusion Matrix: [[14195 870] [1781 13238]] Classification Report: precision recall f1-score support 0 0.89 0.94 0.91 15065 1 0.94 0.88 0.91 15019 accuracy 0.91 macro avg 0.91 weighted avg 0.91 </pre>
--------------------	--

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Model	<p>The Random Forest Classifier performed best with the highest accuracy (95%) and balanced precision and recall, making it both reliable and generalizable. Its ensemble approach reduces overfitting and handles feature interactions better than individual models, leading to consistently strong results across all metrics.</p>